# Parameter Estimation in State Space Models with Multiplicative Noise - Examples

**Møller, Jan Kloppenborg; Carstensen, Niels Jacob; Madsen, Henrik**

*Publication date:*
2009

*Document Version*
Publisher's PDF, also known as Version of record

# Parameter Estimation in State Space Models with Multiplicative Noise - Examples

Jan Kloppenborg Møller[(1,2)] Jacob Carstensen[2], Henrik Madsen[1]
[1] DTU-Informatics
[2] National Environmental Research Institute - NERI

March 24, 2009

# Contents

# Chapter 1

# Introduction

This report is a further elaboration on the results developed in Møller et. al. (2008). The focus is to give some additional examples illustrating the estimation and filtering procedure developed in Møller et. al. (2008) and the reader is referred to this paper for details of the derivation. The model under consideration consists of the system equation

$$\boldsymbol{X}_t = \boldsymbol{\Xi}_t \boldsymbol{A}_t \boldsymbol{X}_{t-1} + \boldsymbol{B}_t \boldsymbol{U}_{t-1}, \tag{1.1}$$

where $\boldsymbol{X}_t \in \mathbb{R}^n$ is the unobserved true state of the system. The input or driver of the system, $\boldsymbol{U}_t \in \mathbb{R}^{n_u}$, can be stochastic or deterministic, but it is assumed to be independent of other variables in the system. If $\boldsymbol{U}_t$ is stochastic, then $\boldsymbol{U}_t$ should be given by its mean $\boldsymbol{\mu}_t^{\boldsymbol{U}}$ and variance $\boldsymbol{\Sigma}_t^{\boldsymbol{U}}$. The system is observed under multiplicative and additive noise

$$\boldsymbol{Y}_t = \boldsymbol{\Lambda}_t \boldsymbol{C}_t \boldsymbol{X}_t + \boldsymbol{\epsilon}_t, \tag{1.2}$$

where $\boldsymbol{Y}_t \in \mathbb{R}^m$ is the observation of the system. In (1.1) and (1.2) it is assumed that the coefficient matrices $\boldsymbol{A}_t \in \mathbb{R}^{n \times n}$, $\boldsymbol{B}_t \in \mathbb{R}^{n \times n_u}$ and $\boldsymbol{C}_t \in \mathbb{R}^{m \times n}$ contain known parametric relationships between states, inputs and observations. $\boldsymbol{\Xi}_t \in \mathbb{R}^{n \times n}$ and $\boldsymbol{\Lambda}_t \in \mathbb{R}^{m \times m}$ are multiplicative noise terms, which are diagonal matrices with diagonal elements given as random vectors $\boldsymbol{\xi}_t \in \mathbb{R}^n$ and $\boldsymbol{\lambda}_t \in \mathbb{R}^m$, with $E[\boldsymbol{\xi}_t] = \boldsymbol{1}$, $E[\boldsymbol{\lambda}_t] = \boldsymbol{1}$, $V[\boldsymbol{\xi}_t] = \boldsymbol{\Sigma}_t^{\boldsymbol{\xi}}$ and $V[\boldsymbol{\lambda}_t] = \boldsymbol{\Sigma}_t^{\boldsymbol{\lambda}}$. The additive noise is a random vector, $\boldsymbol{\epsilon}_t \in \mathbb{R}^m$, characterized by $E[\boldsymbol{\epsilon}_t] = \boldsymbol{\mu}_t^{\boldsymbol{\epsilon}}$ and $V[\boldsymbol{\epsilon}_t] = \boldsymbol{\Sigma}_t^{\boldsymbol{\epsilon}}$.

## 1.1 The Filter Equations

The Kalman filtering procedure for the state space model (1.1)-(1.2) (Møller et. al. (2008)) has the reconstruction

$$\hat{\boldsymbol{X}}_{t|t} = \hat{\boldsymbol{X}}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_{t|t-1}) \tag{1.3}$$

$$\boldsymbol{\Sigma}^{xx}_{t|t} = \boldsymbol{\Sigma}^{xx}_{t|t-1} - \boldsymbol{K}_t(\boldsymbol{\Sigma}^{xy}_{t|t-1})^T. \tag{1.4}$$

with

$$\boldsymbol{K}_t = \boldsymbol{\Sigma}^{xy}_{t|t-1}(\boldsymbol{\Sigma}^{yy}_{t|t-1})^{-1} = \boldsymbol{\Sigma}^{xx}_{t|t-1}\boldsymbol{C}^T_t(\boldsymbol{\Sigma}^{yy}_{t|t-1})^{-1} \tag{1.5}$$

and the predictions

$$\hat{\boldsymbol{X}}_{t+1|t} = \boldsymbol{A}_{t+1}\hat{\boldsymbol{X}}_{t|t} + \boldsymbol{B}_{t+1}\boldsymbol{\mu}^{\boldsymbol{u}}_t \tag{1.6}$$

$$\begin{aligned}\boldsymbol{\Sigma}^{xx}_{t+1|t} =& \boldsymbol{A}_{t+1}\boldsymbol{\Sigma}^{xx}_{t|t}\boldsymbol{A}^T_{t+1} + \boldsymbol{\Sigma}^{\boldsymbol{\xi}}_{t+1} \odot (\boldsymbol{A}_{t+1}\boldsymbol{P}^{xx}_{t|t}\boldsymbol{A}^T_{t+1}) + \\ & \boldsymbol{B}_{t+1}\boldsymbol{\Sigma}^{\boldsymbol{u}}_t\boldsymbol{B}^T_{t+1}\end{aligned} \tag{1.7}$$

$$\hat{\boldsymbol{Y}}_{t+1|t} = \boldsymbol{C}_{t+1}\hat{\boldsymbol{X}}_{t+1|t} + \boldsymbol{\mu}^{\boldsymbol{\epsilon}}_t \tag{1.8}$$

$$\begin{aligned}\boldsymbol{\Sigma}^{yy}_{t+1|t} =& \boldsymbol{C}_{t+1}\boldsymbol{\Sigma}^{xx}_{t+1|t}\boldsymbol{C}^T_{t+1} + \boldsymbol{\Sigma}^{\boldsymbol{\lambda}}_{t+1} \odot (\boldsymbol{C}_{t+1}\boldsymbol{P}^{xx}_{t+1|t}\boldsymbol{C}^T_{t+1}) + \\ & \boldsymbol{\Sigma}^{\boldsymbol{\epsilon}}_{t+1}\end{aligned} \tag{1.9}$$

$$\boldsymbol{\Sigma}^{xy}_{t+1|t} = \boldsymbol{\Sigma}^{xx}_{t+1|t}\boldsymbol{C}^T_{t+1}, \tag{1.10}$$

with

$$\boldsymbol{P}^{xx}_{t+k|t} = \boldsymbol{\Sigma}^{xx}_{t+k|t} + \hat{\boldsymbol{X}}_{t+k|t}\hat{\boldsymbol{X}}^T_{t+k|t}, \tag{1.11}$$

and given initial conditions

$$\hat{\boldsymbol{X}}_{1|0} = \boldsymbol{\mu}_0 \tag{1.12}$$

$$\hat{\boldsymbol{\Sigma}}^{xx}_{1|0} = \boldsymbol{V}_0. \tag{1.13}$$

Positive systems with additive noise is an important subset of the family (1.1)-(1.2), in this case (1.3) is replaced by

$$(\hat{\boldsymbol{X}}_{t|t})_i = \max\{0, (\hat{\boldsymbol{X}}_{t|t-1} + \boldsymbol{\Sigma}^{xy}_{t|t-1}(\boldsymbol{\Sigma}^{yy}_{t|t-1})^{-1}(\boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_{t|t-1}))_i\}. \tag{1.14}$$

The role of this modification is important in the estimation presented in the next section.

## 1.2 Likelihood Estimation

In the estimation procedure we distinguish between two different cases. 1) situations where the sample space for the states and the observations is $\mathbb{R}$ and 2) situations where the sample space for the states and the observations is $\mathbb{R}_0$ only. In both situation we will denote the information up to time $t$ by $\mathcal{Y}_t$. In case 1) we assume that the process is local Gaussian, i.e. $\boldsymbol{Y}_{t+1}|\mathcal{Y}_t$ follows a Gaussian distribution, and the $\log$-likelihood function becomes

$$\log L_G(\boldsymbol{\theta}; \mathcal{Y}_N) = -\frac{1}{2}\sum_{i=1}^{N}\left(\log(\det\boldsymbol{\Sigma}^{yy}_{i|i-1}) + \tilde{\boldsymbol{Y}}^T_{i|i-1}(\boldsymbol{\Sigma}^{yy}_{i|i-1})^{-1}\tilde{\boldsymbol{Y}}_{i|i-1}\right) + \text{constant}, \tag{1.15}$$

where $\tilde{\boldsymbol{Y}}_{i|i-1} = \boldsymbol{Y}_i - \hat{\boldsymbol{Y}}_{i|i-1}$ and the ML estimate of the unknown parameters $\boldsymbol{\theta}$ becomes

$$\hat{\boldsymbol{\theta}}_G = \arg\left\{\max_{\theta} \log L_G(\boldsymbol{\theta}; \mathcal{Y}_N)\right\}. \tag{1.16}$$

In case 2) the assumption is that the process is locally well approximated by a $\log$-normal distribution, i.e. $\boldsymbol{Y}_{t+1}|\mathcal{Y}_t$ follows a $\log$-normal distribution. The filter equation given in Section 1.1 provides the variance matrices which by construction are positive definite. This is however not a sufficient condition for to ensure that the variance matrices are admissible (i.e. it is possible to find a $\log$-normal random variable with this variance) for a $\log$-normal random variable, see Kotz et. al. (2000). The one step prediction variances are therefore transformed in accordance with Møller et. al. (2008), i.e.

$$\tilde{\boldsymbol{\Sigma}}^{zz}_{t+1|t} = \log\left(\tilde{\boldsymbol{\Sigma}}^{yy}_{t+1|t} + \hat{\boldsymbol{Y}}_{t+1|t}\hat{\boldsymbol{Y}}^T_{t+1|t}\right) - \log\left(\hat{\boldsymbol{Y}}_{t+1|t}\hat{\boldsymbol{Y}}^T_{t+1|t})\right) \tag{1.17}$$

$$\hat{\boldsymbol{Z}}_{t+1|t} = \log\left(\hat{\boldsymbol{Y}}_{t+1|t}\right) - \frac{1}{2}\text{diag}\tilde{\boldsymbol{\Sigma}}^{zz}_{t+1|t}, \tag{1.18}$$

with

$$\left(\tilde{\boldsymbol{\Sigma}}^{yy}_{t+1|t}\right)_{i,j} = \hat{Y}_{i,t+1|t}\hat{Y}_{j,t+1|t}\left[\exp\left(\frac{\sigma^{yy}_{ij}}{\sqrt{\sigma^{yy}_{ii}\sigma^{yy}_{jj}}}\sqrt{\log\left(\frac{\sigma^{yy}_{ii} + \hat{Y}^2_{i,t+1|t}}{\hat{Y}^2_{i,t+1|t}}\right)}\times \sqrt{\log\left(\frac{\sigma^{yy}_{jj} + \hat{Y}^2_{j,t+1|t}}{\hat{Y}^2_{j,t+1|t}}\right)}\right) - 1\right], \tag{1.19}$$

where $\sigma_{ij}^{yy} = \left(\boldsymbol{\Sigma}_{t|t+1}^{yy}\right)_{ij}$. This transformation will leave the diagonal elements of $\boldsymbol{\Sigma}_{ij}^{yy}$ unchanged. Under the assumption of local $\log$-normal distributions, $\boldsymbol{Z}_{t+1|t}$ will be normal distributed and the $\log$-likelihood function is

$$\log L_{LN}(\boldsymbol{\theta}; \mathcal{Y}_N) = -\frac{1}{2}\sum_{i=1}^{N}\left(\log(\det\tilde{\boldsymbol{\Sigma}}_{i|i-1}^{zz}) + \tilde{\boldsymbol{Z}}_{i|i-1}^{T}(\tilde{\boldsymbol{\Sigma}}_{i|i-1}^{zz})^{-1}\tilde{\boldsymbol{Z}}_{i|i-1}\right) + \text{constant}, \tag{1.20}$$

the ML estimate of the unknown parameters $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_{LN} = \arg\left\{\max_{\theta}\log L(\boldsymbol{\theta}; \mathcal{Y}_N)\right\}. \tag{1.21}$$

The following sections will present some examples on the use of the filter equations and the maximum likelihood estimation presented above. The example should be seen as an extension of the example presented in Møller et. al. (2008). In the examples where the $\log$-normal likelihood estimation is possible, the performance of the $\log$-normal likelihood and the Gaussian likelihood are compared.

## 1.3 Missing Data

In practical applications missing observations often occur and these have to be taken into account. Missing observations correspond to no information in which case the Kalman gain is zero. I.e. the reconstruction when $\boldsymbol{Y}_t$ is missing becomes

$$\hat{\boldsymbol{X}}_{t|t} = \hat{\boldsymbol{X}}_{t|t-1} \tag{1.22}$$

$$\boldsymbol{\Sigma}_{t|t}^{xx} = \boldsymbol{\Sigma}_{t|t-1}^{xx}. \tag{1.23}$$

This applies if observation $\boldsymbol{Y}_t$ is completely missing, if a data point is only partly missing, then these formulas does not apply. In this case we simply set the information on the missing observation equal to zero, which is equivalent to infinite variance on the observation. The variance of the observation is given by (equation (1.9))

$$\boldsymbol{\Sigma}_{t+1|t}^{yy} = \boldsymbol{C}_{t+1}\boldsymbol{\Sigma}_{t+1|t}^{xx}\boldsymbol{C}_{t+1}^{T} + \boldsymbol{\Sigma}_{t+1}^{\boldsymbol{\lambda}} \odot (\boldsymbol{C}_{t+1}\boldsymbol{P}_{t+1|t}^{xx}\boldsymbol{C}_{t+1}^{T}) + \boldsymbol{\Sigma}_{t+1}^{\boldsymbol{\epsilon}}, \tag{1.24}$$

with $\boldsymbol{P}_{t+1|t}^{xx}$ given by (1.11). The variance given by (1.24) can be controlled by the variance of the additive and the multiplicative noises. In the model under consideration the multiplicative variance is not suited for the purpose of controlling the variance, since an observation prediction equal to zero will produce variance corresponding to $\boldsymbol{\Sigma}^{\epsilon}_{t+1}$. The information should therefore be controlled by the additive variance, i.e. if the observation $j$ at time $t$, $Y_{j,t}$, is missing then we set $V[\epsilon_{j,t}] = \infty$. I.e.

$$\left(\boldsymbol{\Sigma}^{\epsilon}_{t+1}\right)_{j,j} = \infty, \quad \text{if} \quad Y_{j,t} \quad \text{missing}, \tag{1.25}$$

if an observation is completely missing then the diagonal elements of $\boldsymbol{\Sigma}^{yy}_{t|t-1}$ will be equal to $\infty$ and $\boldsymbol{K}_t$ becomes the zero matrix and the reconstruction is (1.22)-(1.23). If an observation is only partly missing then (1.25) will set the effect of the missing observation equal to zero. With this the filter equations can follow the updating algorithms given in Section 1.1.

### The Likelihood Function

The approach above is useful for the filtering problem. For the maximum likelihood estimation problem we need to remove the missing part of the observation from the estimation procedure.

If an observation is completely missing then it is simply ignored in the likelihood estimation. If an observation is partly missing then let $\mathcal{J}_t$ denote the index non-missing part of the observation at time $t$. E.g. if a full observation at time $t$ consists of $\boldsymbol{Y}_t = (Y_{1,t}, Y_{2,t}, Y_{3,t})$, and $Y_{2,t}$ is missing, then $\mathcal{J}_t = \{1, 3\}$. Now set

$$\boldsymbol{C}'_t = \boldsymbol{I}_{\mathcal{J}_t,:}, \tag{1.26}$$

where $\boldsymbol{I}$ refer to the identity matrix of dimension equal to the observation variance ($\mathbb{R}^{m\times m}$) and $\boldsymbol{I}_{\mathcal{J}_t}$ refer to the $\mathcal{J}_t$ rows and all columns. Define a new observation by

$$\boldsymbol{Y}'_t = \boldsymbol{C}'_t \boldsymbol{Y}_t, \tag{1.27}$$

with the variance

$$\boldsymbol{\Sigma}'^{yy}_{t+1|t} = \boldsymbol{C}'_t \boldsymbol{\Sigma}^{yy}_{t+1|t} \boldsymbol{C}'^{T}_t. \tag{1.28}$$

Now $\boldsymbol{Y}'_t$ and $\boldsymbol{\Sigma}'^{yy}_{t+1|t}$ is simply used in the likelihood estimation as described in Section 1.2, with the remark that observations, that are completely missing, are simply ignored, i.e. left out in the summations (1.15) and (1.20).

# Chapter 2

# Examples

This chapter will present some examples of applications of the filter and the likelihood presented in Section 1.1 and 1.2 to illustrate the effect of missing observations.

## 2.1 Example 1

The first example we consider is the following state space model

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \xi_{1,t} & 0 \\ 0 & \xi_{2,t} \end{bmatrix} \begin{bmatrix} 0.3 & 0.6 \\ 0.2 & 0.7 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} U_t \tag{2.1}$$

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} \lambda_{1,t} & 0 \\ 0 & \lambda_{2,t} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}, \tag{2.2}$$

with $\boldsymbol{\xi}_t$ and $\boldsymbol{\lambda}_t$ $\log$-normal distributed random variables with covariances

$$\boldsymbol{\Sigma}_t^{\boldsymbol{\xi}} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.2 \end{bmatrix}, \quad \boldsymbol{\Sigma}_t^{\boldsymbol{\lambda}} = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.7 \end{bmatrix}, \tag{2.3}$$

and expectation equal to the identity matrix. $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are independent iid sequences of exponentially distributed random variables with mean $0.01$. The initial values are set at

$$\hat{\boldsymbol{X}}_{1|0} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{1|0}^{xx} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.4}$$

The largest eigenvalue of $\boldsymbol{A}$ is 0.9, which imply that the memory of the system is longer than in the example presented in Møller et. al. (2008). Figure 2.1

give a clear idea of how the system noise is scale dependent, the long memory in comparison with the example in Møller et. al. (2008) is also evident. The dynamics when the state is close to zero is however not clear from a plot like Figure 2.1. Figure 2.2 gives insight of the dynamics when the state is close to zero. The scale dependency of the noise is clear from Figure 2.2, where the width of the confidence band is almost constant over the scales in the $\log$-domain, implying that it scales with the state in the original domain. It is however important to emphasize that the state estimation and the simulation have been done in the original domain.

The confidence bands in Figure 2.1 and 2.2 are calculated as the unconditional (on the other state) confidence region under the assumption that the states are $\log$-normal distributed. The confidence regions for the two-dimensional process cannot be shown in figures like Figures 2.1 and 2.2. These has however been calculated, and for the shown simulation 95.4% of the states are inside confidence region for the reconstruction and 96.9% of the observations are inside the confidence region for the one step prediction. The comparison was done in the log domain as the parametric form of the confidence regions for $\log$-normal variables is quite complicated, as can be seen in Appendix A.

The multiplicative standard deviation (see Limpert et. al. (2001), and Møller et. al. (2008) for the filter version) in Figure 2.3 give a good picture of when the uncertainty in the $\log$-domain is large. The multiplicative standard deviation is relatively constant over the simulation, some larger values is however seen when the states of the system are small. This is also the situations where the influence of the additive noise will take over in the simulations.

For the estimation part we assume that $\boldsymbol{C}$ is known and that the correlation structure of the noise is known. We also assume that the input $U_t$ and $\boldsymbol{B}$ is known. With this the vector of unknown parameters becomes

$$\boldsymbol{\theta} = [a_{1,1}, a_{1,2}, a_{2,1}, a_{2,2}, \sigma^2_{\xi_1}, \sigma^2_{\xi_2}, \sigma^2_{\lambda_1}, \sigma^2_{\lambda_2}, \mu_{\epsilon_1}, \mu_{\epsilon_2}]. \tag{2.5}$$

since all parameters have to be greater than zero and therefore we set

$$\boldsymbol{\psi} = \log(\boldsymbol{\theta}), \tag{2.6}$$

and optimize the likelihood with respect to $\boldsymbol{\psi}$, in this way we can estimate from the real line and avoid boundary problems in the optimization procedure. We use the standard optimizer "`optim`" in "R" (see `www.r-project.org`) with the method "`Nelder-Mead`". The parameters are estimated with the local Gaussian and the local $\log$-normal assumption. Figure 2.4 shows that the local $\log$-normal assumption perform better than the local Gaussian assumption in all the parameter with the possible exception of $\mu_{\epsilon_1}$ and $\mu_{\epsilon_2}$.

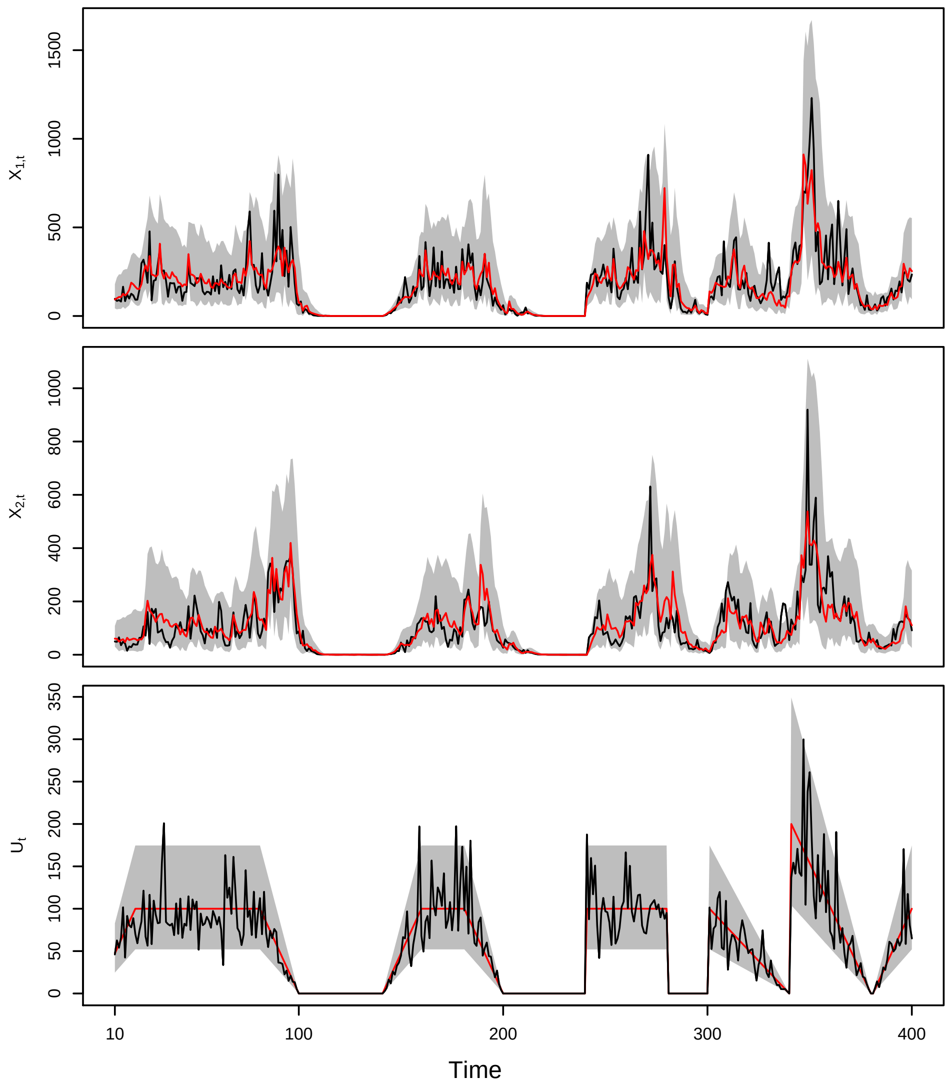


Figure 2.1: The true state $\boldsymbol{X}_t = (X_{1,t}, X_{2,t})$ of the system (black), the state reconstruction $\hat{\boldsymbol{X}}_{t|t}$ of the system (red), with 95% confidence intervals (grey area), forced by the stochastic input $U_t$ (black), with mean $\mu_t^U$ (red) and 95% confidence interval for input (grey area). The first 10 data-points are considered as transients and are therefore not shown.

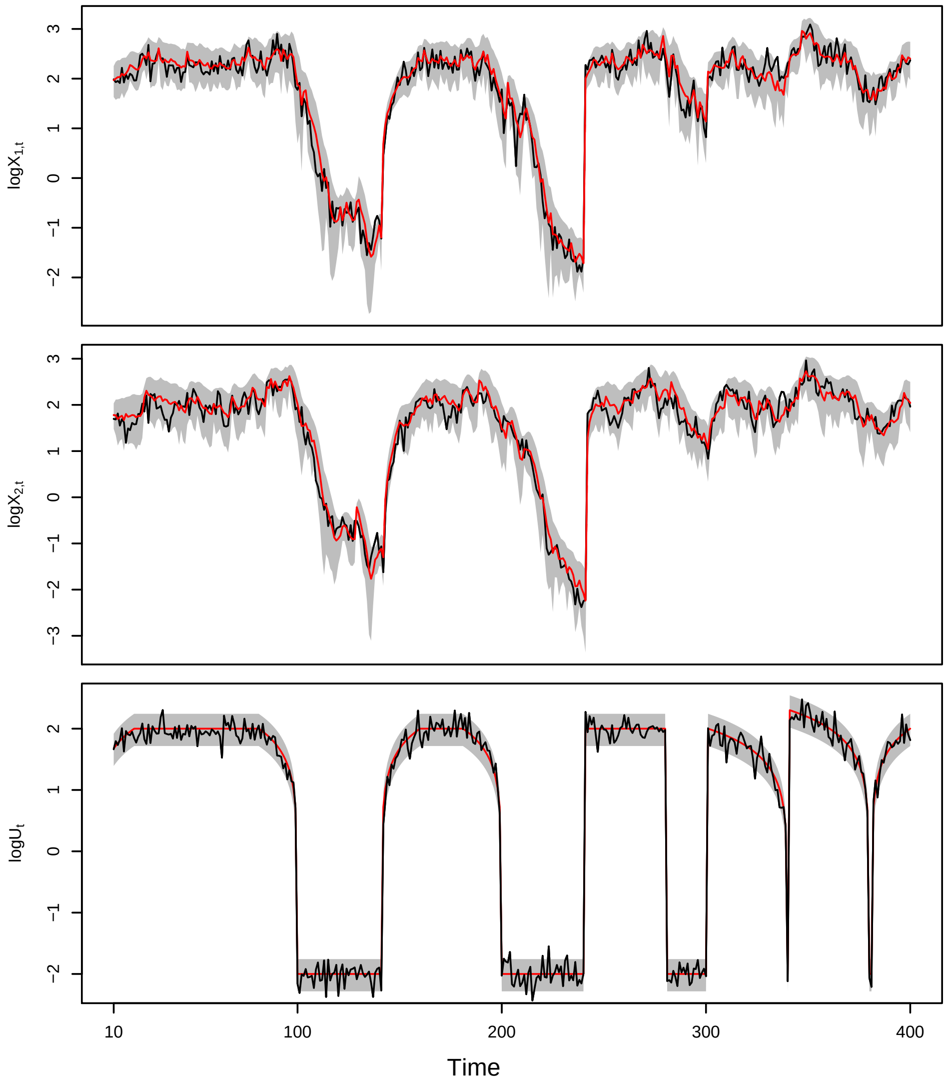


Figure 2.2: The logarithm of the true state $\log(\boldsymbol{X}_t) = (\log X_{1,t}, \log X_{2,t})$ of the system (black), the logarithm of the state reconstruction $\log(\hat{\boldsymbol{X}}_{t|t})$ of the system (red), with 95% confidence intervals (grey area), the logarithm of the stochastic input $\log(U_t)$ (black), $\log \mu_t^U$ (red) and the 95% confidence interval for input (grey area). The first 10 data-points are considered as transients and are therefore not shown.

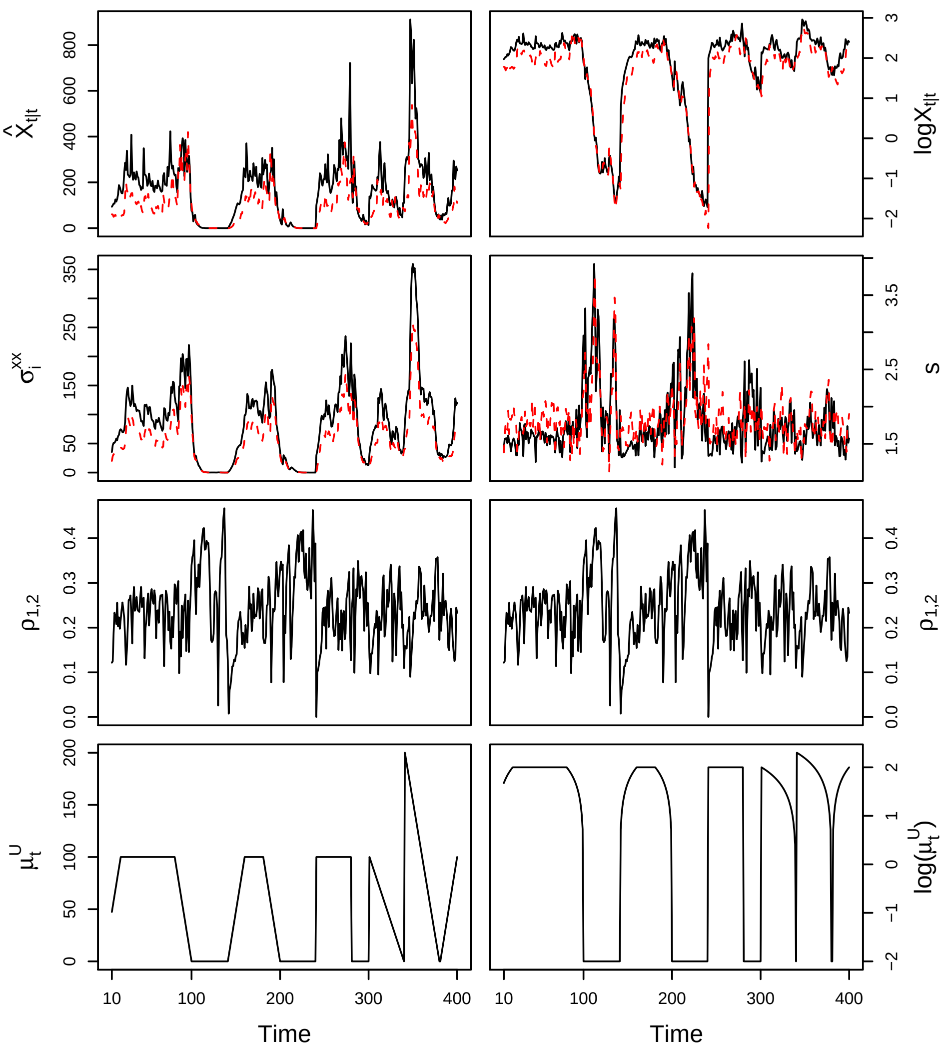


Figure 2.3: Left panel: the states (State 1, black and State 2, red), the standard deviation of the states, the correlation between the states and input on the original scale. Right panel: the logarithm of the states, the "multiplicative" standard deviation, the correlation and the logarithm of the input.

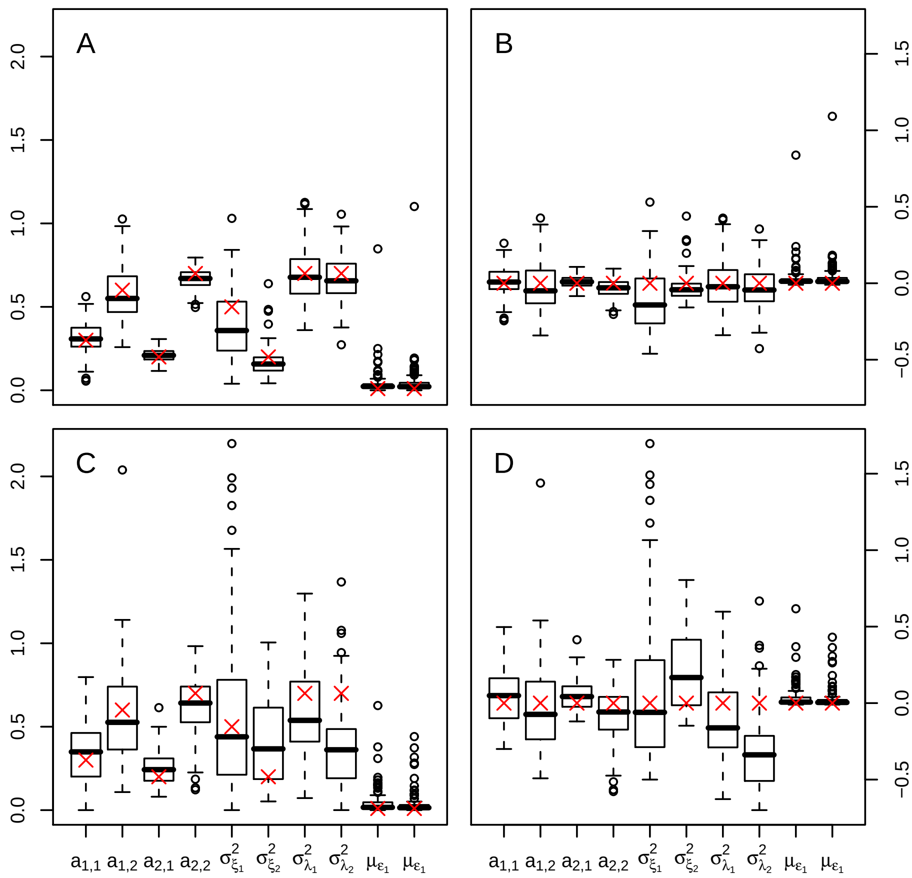


Figure 2.4: Box plots of estimated parameters ($\hat{\boldsymbol{\theta}}_{LN}$) based on 100 simulated processes of (2.1)-(2.3). The boxes show the IQR (Inter Quartile Range), the median is marked with the bold line in the boxes. The whiskers are placed at the most extreme values in the interval [1st quartile- $1.5{\cdot}$IQR, 3rd quartile+$1.5{\cdot}$IQR], while more deviating parameter estimates are marked as outliers and plotted with black circles. Left panel shows the actual parameter estimates, while the right panel shows the deviation from the true parameter values. A and B are based on the local $\log$-normal estimation procedure whereas C and D are based on the local Gaussian estimation procedure.

The two estimation procedures are tested against each other with a $F$-test. In order to do so define the loss-function

$$S(\hat{\psi}_{:,j}; M) = \sum_{i=1}^{M} (\hat{\psi}_{i,j} - \log(\theta_j)), \tag{2.7}$$

where $\hat{\psi}_{:,j}$ refer to a vector of estimates of parameter $j$ based independent realizations of the process and $\hat{\psi}_{i,j}$ refer to the $i$'th estimate of parameter $j$. $S(\hat{\psi}_{:,j}; M)$ is $\chi^2$ distributed under the assumption that $\hat{\psi}_{i,j}$ is Gaussian distributed and we can compare the ratio between the loss-functions for the two estimation procedure with an $F$-distribution. Table 2.1 shows that the $\log$-normal procedure performs better than the Gaussian procedure on a 0.001 level for all the parameters in the model in the considered example.

As was shown in Møller et. al. (2008) the conclusion will depend on the skewness of the multiplicative noise and in the example considered here the variance of the multiplicative noise is large and therefore the skewness is high.

Table 2.1: Loss-functions for parameter estimates based on the two estimation procedures ($\log$-normal and Gausian), with $M = 100$ independent realizations of the process described in Example 1 the ratio between the loss-function and an $F$-test of the hypothesis that the estimations are equally good.

| | $S(\hat{\boldsymbol{\psi}}_{LN;:,j}, M)$ | $S(\hat{\boldsymbol{\psi}}_{G;:,j}, M)$ | $F = \frac{S(\hat{\boldsymbol{\Psi}}_{G;:,j},M)}{S(\hat{\boldsymbol{\Psi}}_{LN;:,j},M)}$ | $P\{x < F\}$ |
|---|---|---|---|---|
| $a_{1,1}$ | 1.73e−01 | 1.84e+00 | 10.59 | 1.000 |
| $a_{1,2}$ | 8.51e−02 | 3.43e−01 | 4.04 | 1.000 |
| $a_{2,1}$ | 3.71e−02 | 1.97e−01 | 5.30 | 1.000 |
| $a_{2,2}$ | 1.15e−02 | 1.61e−01 | 13.98 | 1.000 |
| $\sigma^2_{\xi_1}$ | 6.19e−01 | 4.00e+00 | 6.46 | 1.000 |
| $\sigma^2_{\xi_2}$ | 2.70e−01 | 8.37e−01 | 3.09 | 1.000 |
| $\sigma^2_{\lambda_1}$ | 5.11e−02 | 3.99e−01 | 7.81 | 1.000 |
| $\sigma^2_{\lambda_2}$ | 5.06e−02 | 3.25e+00 | 64.24 | 1.000 |
| $\mu_{\epsilon_1}$ | 2.78e+00 | 2.70e+01 | 9.74 | 1.000 |
| $\mu_{\epsilon_2}$ | 2.91e+00 | 1.26e+01 | 4.34 | 1.000 |

## 2.2 Example 2

As a second example we consider the following state space model

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \xi_{1,t} & 0 \\ 0 & \xi_{2,t} \end{bmatrix} \begin{bmatrix} 0.3 & 0.6 \\ 0.2 & 0.7 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} U_t \tag{2.8}$$

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} \lambda_{1,t} & 0 \\ 0 & \lambda_{2,t} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}, \tag{2.9}$$

with $\boldsymbol{\xi}_t$ and $\boldsymbol{\lambda}_t$ $\log$-normal distributed random variables with covariances

$$\boldsymbol{\Sigma}_t^{\boldsymbol{\xi}} = \begin{bmatrix} 0.5 & 0.28 \\ 0.28 & 0.2 \end{bmatrix}, \quad \boldsymbol{\Sigma}_t^{\boldsymbol{\lambda}} = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.7 \end{bmatrix}, \tag{2.10}$$

respectively, and expectation equal to the identity matrix, the covariance between $\xi_1$ and $\xi_2$ correspond to a correlation equal to 0.9. $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are independent i.i.d. sequences of exponentially distributed random variables with mean $0.01$. The initial values are set at

$$\hat{\boldsymbol{X}}_{1|0} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{1|0}^{xx} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.11}$$

This system is identical to the system presented in Example 1, except that the correlation between the multiplicative system noise is different from zero. The impact from introducing this correlation is clear from Figure 2.5, where it is seen that the correlation between the reconstruction of the states becomes very large compared to Example 1. In addition some abrupt changes in the correlation appears when the loading increases dramatically (in the $\log$-domain), these are due to the dominance of the input variance around these changes in the input. The confidence regions are also calculated in this example and the number of true states in the confidence region for the reconstruction of states is 94.9% and the number is 96.2% for the one step prediction of the observations.

In this example we will also compare the estimation procedures, and as in Example 1 the parameters are transformed before the actual optimization takes place. The large correlation between the multiplicative noise terms may course numerical problems for the estimation procedure. These problems stem from the estimation procedure setting some parameters equal to zero. The transformation from Example 1 is therefore changed to

$$\psi_j = \log(\theta_j + 10^{-4}), \tag{2.12}$$

this transformation is carried out for all the parameter except for the correlation. The correlation is bounded by the interval $[\rho_{min}, \rho_{max}]$, which can be found by inserting $\rho = \pm 1$ in equation (A.3). By using a logit transformation the correlation is bounded within this interval. The interval depends on the variances, and the transformation will therefore change as the optimization procedure progresses.

Figure 2.6 show that the estimation $\log$-normal estimation procedure does a better job on the estimation than the Gaussian estimation procedure. It is also seen that the quality of the estimation seems to be good without large bias problems. Table 2.2 confirms the conclusions from Figure 2.6 and yields convincing statistics on the problem.

Table 2.2: Loss-functions for parameter estimates based on the two estimation procedures ($\log$-normal and Gausian), with $M = 100$ independent realizations of the process described in Example 2 the ratio between the loss-function and an $F$-test of the hypothesis that the estimations are equally good.

| | $S(\hat{\mathbf{\Psi}}_{LN}, M_1)_j$ | $S(\hat{\mathbf{\Psi}}_G, M_2)_j$ | $F = \frac{S(\hat{\mathbf{\Psi}}_G, M_2)_j}{S(\hat{\mathbf{\Psi}}_{LN}, M_1)_j}$ | $P\{x < F\}$ |
|---|---|---|---|---|
| $a_{1,1}$ | 1.57e−01 | 8.04e−01 | 5.11 | 1.000 |
| $a_{1,2}$ | 9.22e−02 | 6.45e−01 | 6.99 | 1.000 |
| $a_{2,1}$ | 4.12e−02 | 1.22e−01 | 2.96 | 1.000 |
| $a_{2,2}$ | 1.66e−02 | 8.37e−02 | 5.04 | 1.000 |
| $\sigma^2_{\xi_1}$ | 2.25e−01 | 9.47e−01 | 4.21 | 1.000 |
| $\rho_{\xi_{1,2}}$ | 1.44e+01 | 6.72e+01 | 4.68 | 1.000 |
| $\sigma^2_{\xi_2}$ | 2.67e−01 | 6.40e−01 | 2.40 | 1.000 |
| $\sigma^2_{\lambda_1}$ | 6.09e−02 | 4.70e−01 | 7.72 | 1.000 |
| $\sigma^2_{\lambda_2}$ | 1.26e−01 | 1.21e+00 | 9.66 | 1.000 |
| $\mu_{\epsilon_1}$ | 1.33e+00 | 2.47e+01 | 18.57 | 1.000 |
| $\mu_{\epsilon_2}$ | 1.45e+00 | 1.56e+01 | 10.79 | 1.000 |

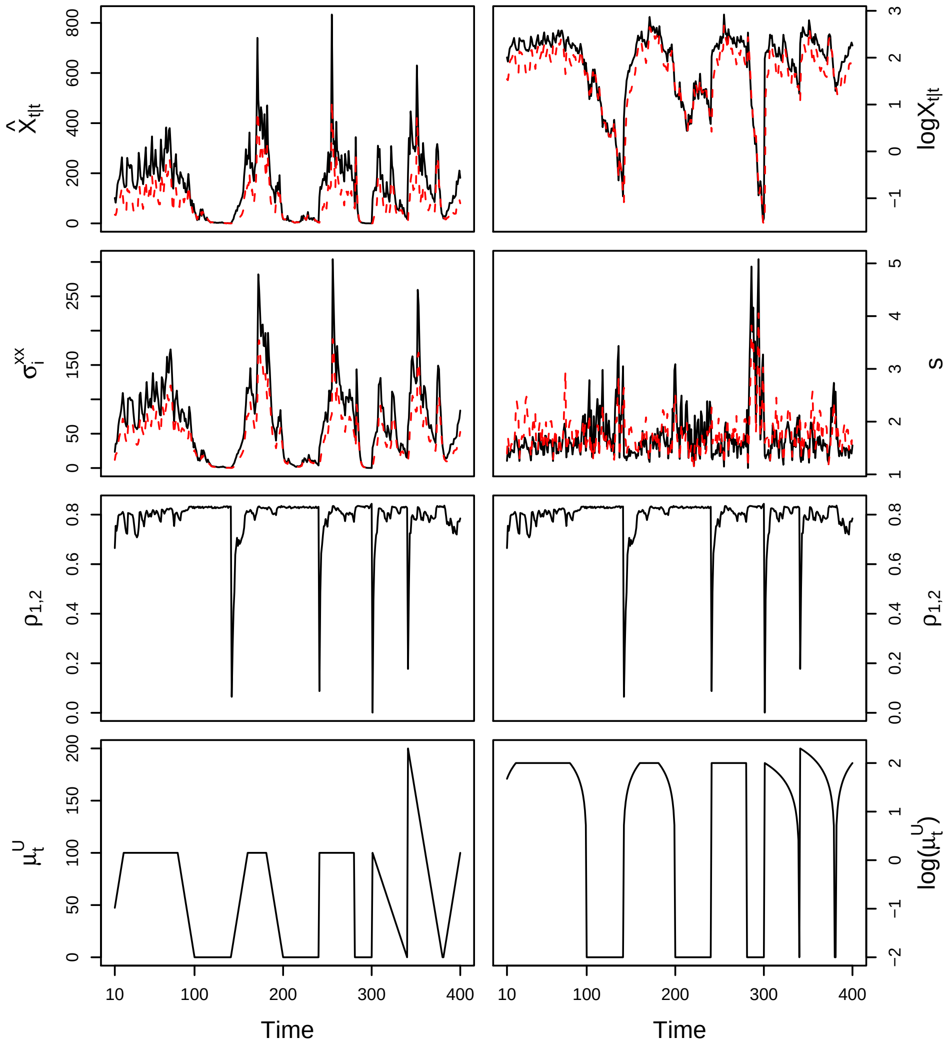


Figure 2.5: Left panel: the states (State 1, black and State 2, red), the standard deviation of the states, the correlation between the states and input on the original scale. Right panel: the logarithm of the states, the "multiplicative" standard deviation, the correlation and the logarithm of the input.

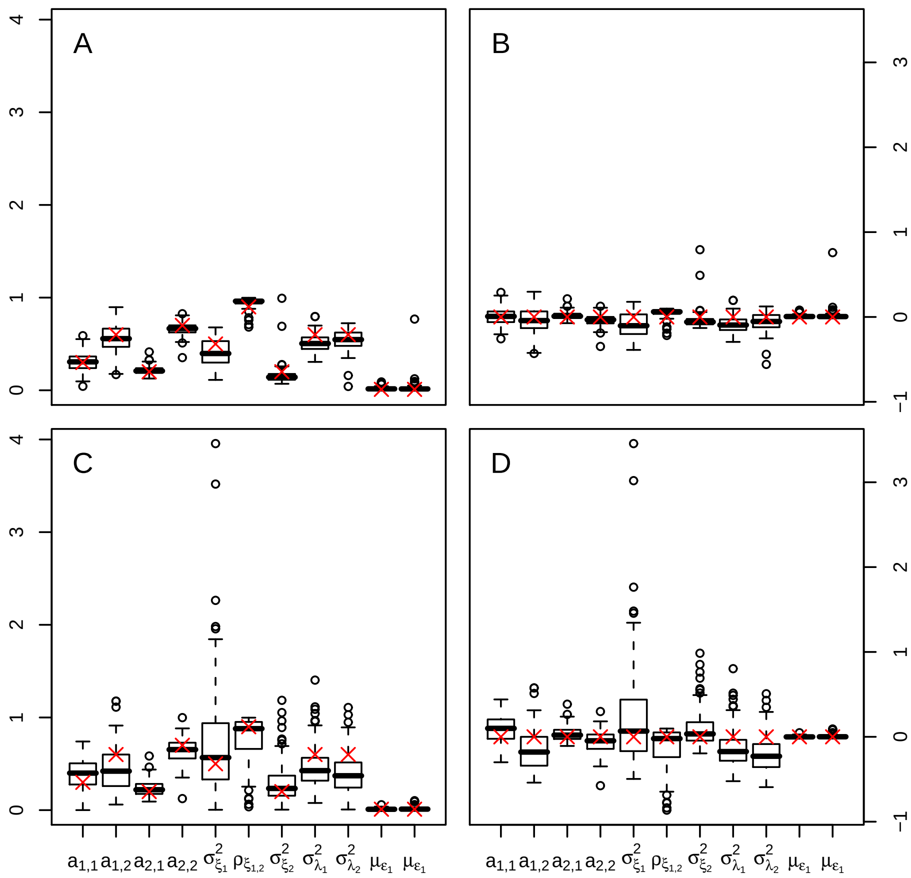


Figure 2.6: Box plots of estimated parameters ($\hat{\boldsymbol{\theta}}_{LN}$) based on 100 simulated processes of (2.8)-(2.10). The boxes show the IQR (Inter Quartile Range), the median is marked with the bold line in the boxes. The whiskers are placed at the most extreme values in the interval [1st quartile- $1.5{\cdot}$IQR, 3rd quartile+$1.5{\cdot}$IQR], while more deviating parameter estimates are marked as outliers and plotted with black circles. Left panel shows the actual parameter estimates, while the right panel shows the deviation from the true parameter values. A and B are based on the local $\log$-normal estimation procedure whereas C and D are based on the local Gaussian estimation procedure.

## 2.3 Example 3

As a third example of the filtering and estimation problem we consider the following state space model

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \xi_{1,t} & 0 \\ 0 & \xi_{2,t} \end{bmatrix} \begin{bmatrix} 0.5 & -0.6 \\ 0.3 & 0.8 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} U_t, \tag{2.13}$$

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} \lambda_{1,t} & 0 \\ 0 & \lambda_{2,t} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}, \tag{2.14}$$

with $\boldsymbol{\xi}_t$ and $\boldsymbol{\lambda}_t$ $\log$-normal distributed random variables with covariances

$$\boldsymbol{\Sigma}_t^{\boldsymbol{\xi}} = \begin{bmatrix} 0.2 & 0.25 \\ 0.25 & 0.4 \end{bmatrix}, \quad \boldsymbol{\Sigma}_t^{\boldsymbol{\lambda}} = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.7 \end{bmatrix}, \tag{2.15}$$

respectively and expectation equal to the identity matrix, the covariance between $\xi_1$ and $\xi_2$ correspond to a correlation equal to 0.9. $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are independent i.i.d. sequences of exponentially distributed random variables with mean $0.01$. The initial values are set at

$$\hat{\boldsymbol{X}}_{1|0} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{1|0}^{xx} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.16}$$

The model considered here can procedure negative values, and as stated in Møller et. al. (2008) the filter is capable of handling this. The $\log$-normal estimation procedure is however not capable of this and it is therefore not possible to compare the two estimation procedures. We will however consider the result of the local Gaussian assumption.

The evolution of the states can be seen in Figure 2.7, the states do take negative values, but the scale dependency of the noise is still evident from the plot. The states cannot be $\log$-normal distributed in this case, and the confidence intervals are therefore constructed using a Gaussian assumption on the reconstruction, the intervals are confidence intervals in the unconditional distribution. The number of true states inside the 95% confidence interval are 95% and 96.3% for the reconstruction and the one step prediction of the observations respectively.

When the process is guaranteed to be positive we can take the logarithm to visualize the dynamics of the process when close to zero. This is not possible when the process can take negative values. Instead we use the following transformation

$$g(x) = \begin{cases} k \cdot (-\log(-x) + \log(k) - 1) & \text{for} \quad x < -k \\ x & \text{for} \quad -k \leq x \leq k \\ k \cdot (\log(x) - \log(k) + 1) & \text{for} \quad k < x \end{cases}, \tag{2.17}$$

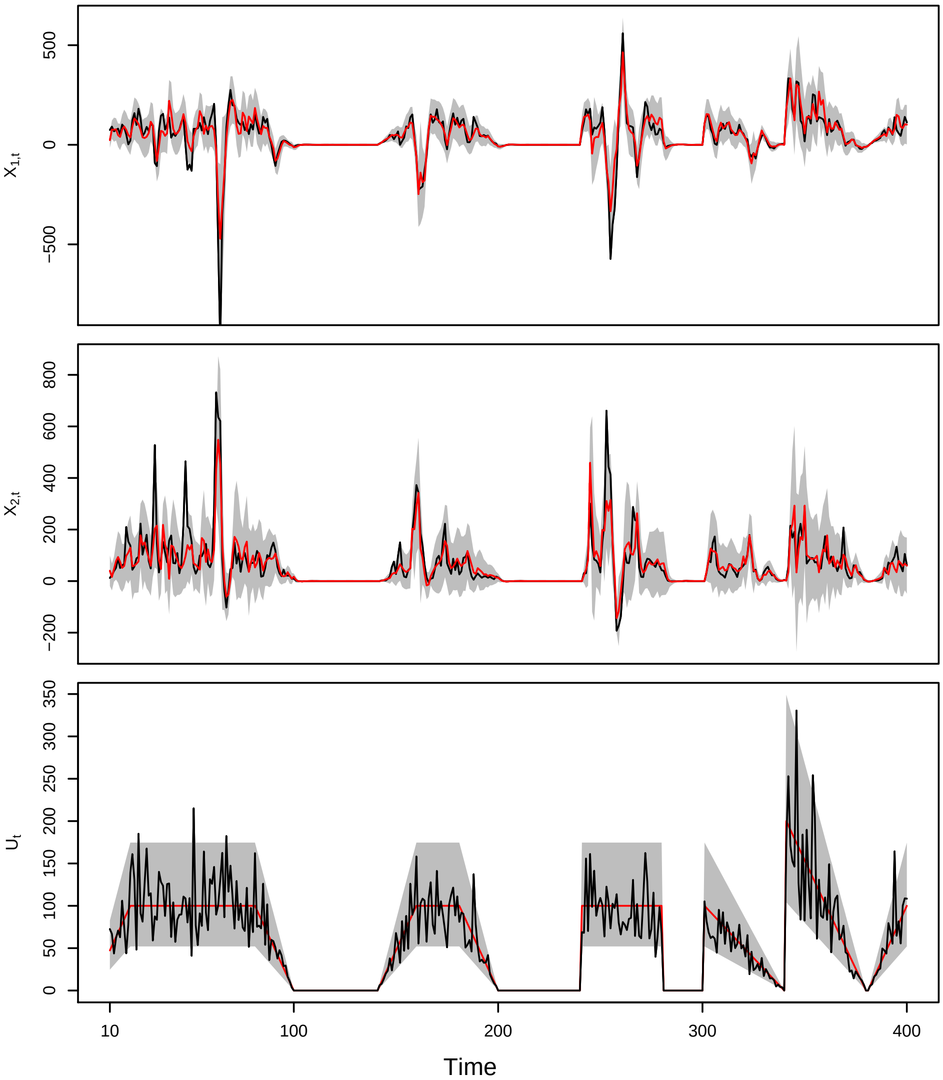


Figure 2.7: The true state $\boldsymbol{X}_t = (X_{1,t}, X_{2,t})$ of the system (black), the state reconstruction $\hat{\boldsymbol{X}}_{t|t}$ of the system (red), with 95% confidence intervals (grey area), forced by the stochastic input $U_t$ (black), with mean $\mu_t^U$ (red) and 95% confidence interval for input (grey area). The first 10 data-points are considered as transients and are therefore not shown.

with $k = 10^{-3}$. This transformation allows us to view the dynamics of the process on different scales on the same plots (Figure 2.8). Even though the plot gives some insight on the dynamics over different scales, the figure is less clear than the $\log$ plots in the previous two examples. The multiplicative standard deviation shows some very large values compared to what has been seen in the previous two example.

Estimation of the parameters cannot be done under the $\log$-normal assumption and we have therefore used the local Gaussian assumption only. As for the previous examples we have transformed the parameters before the estimation was carried out. The parameters in $\boldsymbol{A}$ are not transformed, since these are now allowed to take values from the real line. Variance parameters are considered as positive and are therefore subject to the transformation (2.12). This transformation is also applied to the mean value parameters $\mu_{\epsilon_1}$ and $\mu_{\epsilon_2}$. Finally, the correlation coefficient in $\boldsymbol{\Sigma}_t^{\boldsymbol{\xi}}$ is transformed with the same logit transformation as in Example 2. From Figure 2.9 it is clear that the quality of the estimates are not as good as the results from the $\log$-normal procedure in Example 1 and 2. There are biases and many outliers, especially for the variance (and correlation) parameters.

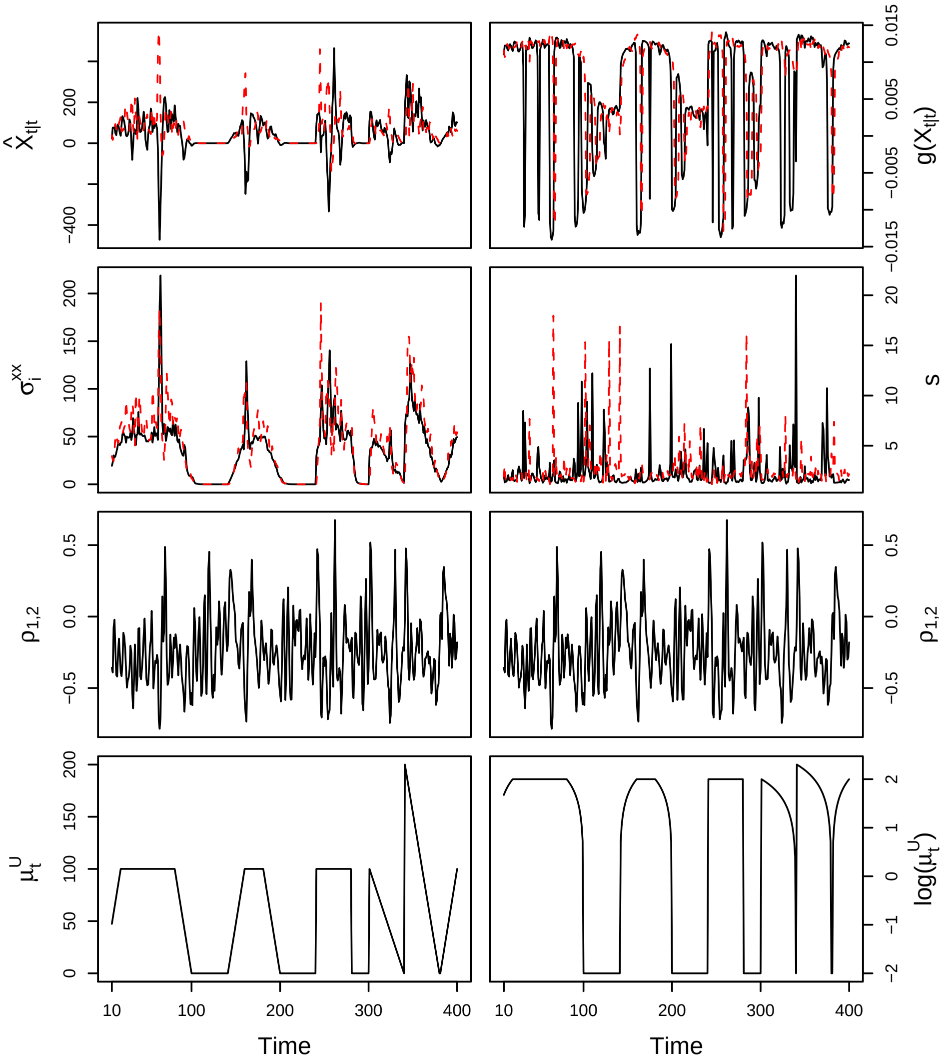


Figure 2.8: Left panel: the states (State 1, black and State 2, red), the standard deviation of the states, the correlation between the states and input on the original scale. Right panel: the transformed states, the "multiplicative" standard deviation, the correlation and the logarithm of the input.

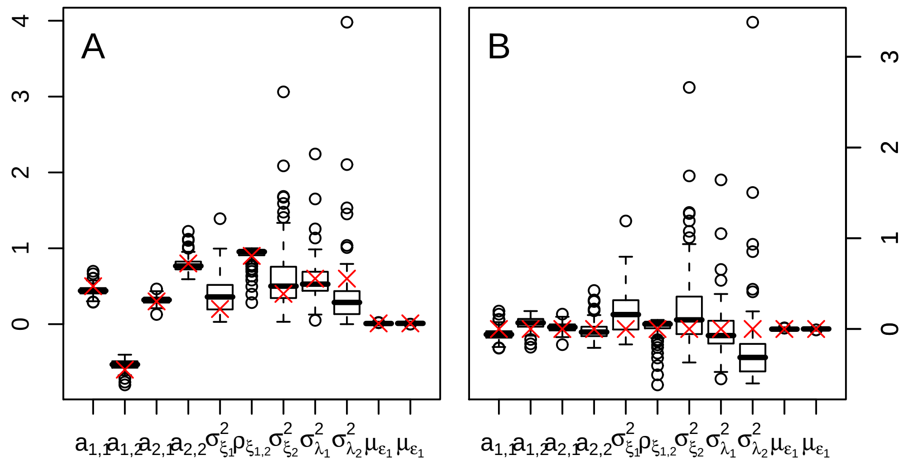

Figure 2.9: Box plots of estimated parameters ($\hat{\boldsymbol{\theta}}$) based on 100 simulated processes of (2.13)-(2.15). The boxes show the IQR (Inter Quartile Range), the median is marked with the bold line in the boxes. The whiskers are placed at the most extreme values in the interval [1st quartile- $1.5 \cdot$IQR, 3rd quartile+$1.5 \cdot$IQR], while more deviating parameter estimates are marked as outliers and plotted with black circles. Left panel shows the actual parameter estimates, while the right panel shows the deviation from the true parameter values. The estimates are based on the local Gaussian estimation procedure.

## 2.4 Example 4 (Missing Data)

As a final example we consider the following state space model

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \xi_{1,t} & 0 \\ 0 & \xi_{2,t} \end{bmatrix} \begin{bmatrix} 0.3 & 0.6 \\ 0.2 & 0.7 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} U_t \tag{2.18}$$

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} \lambda_{1,t} & 0 \\ 0 & \lambda_{2,t} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}, \tag{2.19}$$

with $\boldsymbol{\xi}_t$ and $\boldsymbol{\lambda}_t$ $\log$-normal distributed random variables with covariances

$$\boldsymbol{\Sigma}_t^{\boldsymbol{\xi}} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.2 \end{bmatrix}, \quad \boldsymbol{\Sigma}_t^{\boldsymbol{\lambda}} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}, \tag{2.20}$$

respectively and expectation equal to the identity matrix. $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are independent i.i.d. sequences of exponentially distributed random variables with mean $0.01$. The initial values are set at

$$\hat{\boldsymbol{X}}_{1|0} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{1|0}^{xx} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.21}$$

The model considered here is similar to the one presented in Section 2.1, but with lower observation variance. The model is considered with a loading drawn from a $log$-normal distribution with mean equal to 100 and variance equal to 1000. The effect of missing or partly missing observations on the reconstruction of the state is investigated by letting either $Y_{1,t}$, $Y_{2,t}$ or both be missing. From Figure 2.10 it is seen that the effect of observations from one state being missing does not affect the reconstruction of the other state significantly, and further that the observation of one state gives some information on the other state. The effect on the variance (Figure 2.11) is more clear when observations are partly missing only. The variance and the multiplicative standard deviation is seen to be higher for both states as is the correlation between the states.

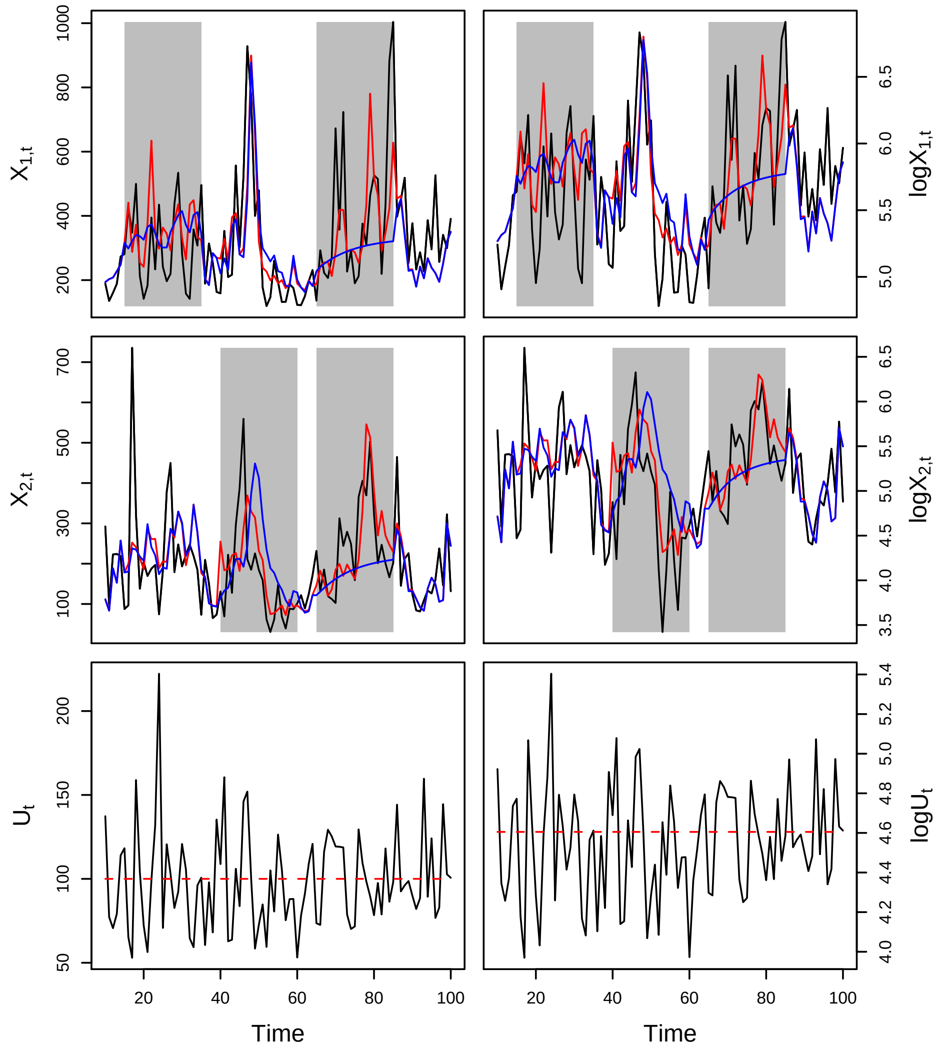


Figure 2.10: The true state of the system (black line), the state reconstruction when all observations are available (red line), and the state reconstruction when the observations marked by the grey area are missing. Further the input for the system is shown. Left panel is in the original domain while the right panel is in the log-domain.

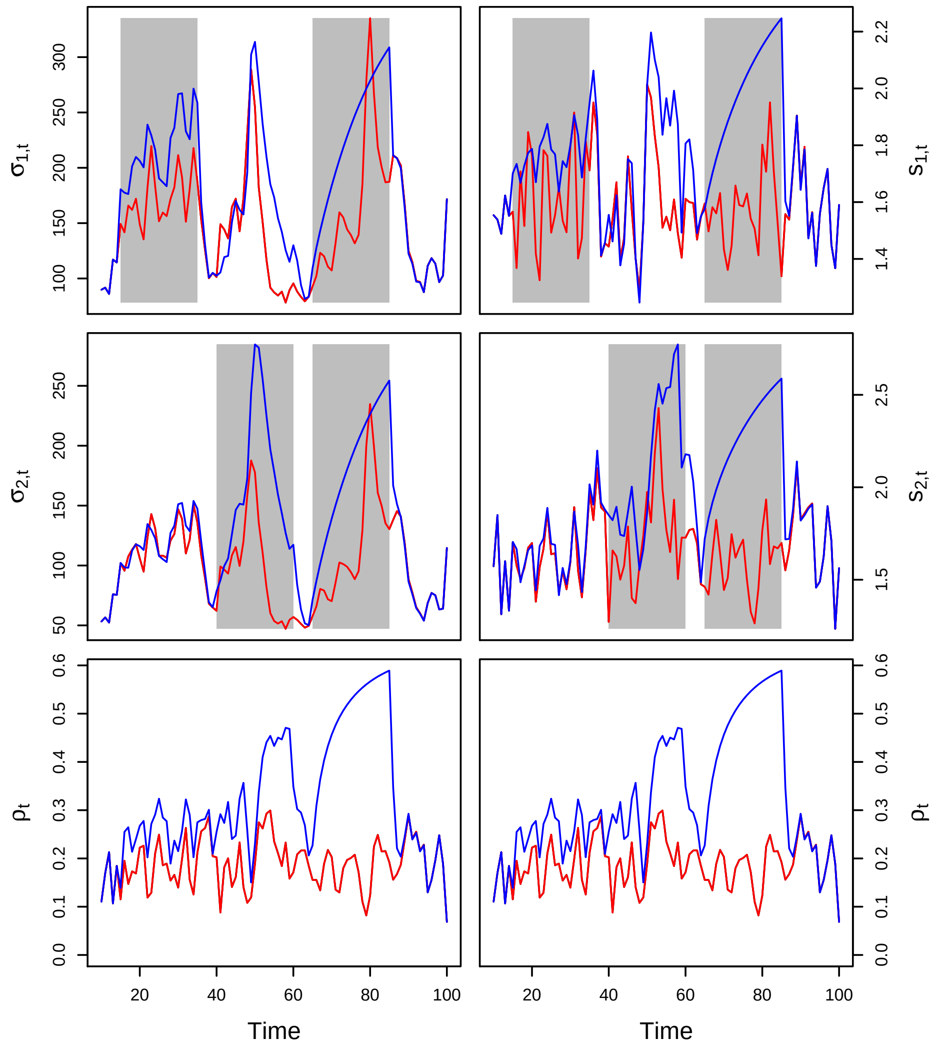


Figure 2.11: Left panel: The standard deviation of the reconstruction when no data is missing (red line), when data indicated by the grey area is missing (blue line) and the cross-correlation of the reconstructions. Right panel: The multiplicative standard deviation of the reconstruction when no data is missing (red line), when data indicated by the grey area is missing (blue line) and the cross-correlation of the reconstructions.

# Chapter 3

# Conclusion and Acknowledgment

## 3.1 Conclusion

The filter presented in Chapter 1 has been applied for four different examples where it has proven to work well. The filter has further been applied as the basis for estimation of parameters in processes with multiplicative noise, the estimation based on $\log$-normal assumption has proven work well and to perform better than the traditional assumption of Gaussian one step transition probabilities.

The results presented here should be seen as further simulation studies in relation to the results presented in Møller et. al. (2008), and as such the conclusions given therein are supported in the present work. The application of the filter to a process that takes values on the real line (as oppose to the positive axes), is not presented in Møller et. al. (2008). The results for this part of the study is that the filter seems to give nice results, but that the local Gaussian assumption is not very satisfactory for the estimation procedure.

The present work further elaborate on how to deal with missing data.

## 3.2 Acknowledgment

This study is a contribution to the EU-funded project Thresholds (GOCE-003933).

# Appendix A

# Log-normal Confidence Regions

The purpose of this section is to visualize how confidence regions for Gaussian random variables translates into confidence regions for the corresponding $\log$-normal random variable.

If $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a multivariate Gaussian random variable, then $\boldsymbol{Y} = e^{\boldsymbol{X}}$ (the exponential is to be taken element-vise) is a multivariate $\log$-normal random variable, we denote this by $\boldsymbol{Y} \sim LN(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$, with (see Kotz et. al. (2000))

$$\tilde{\mu}_i = e^{\mu_i + \frac{1}{2}\sigma_{ii}}, \tag{A.1}$$

$$\tilde{\sigma}_{ij} = (e^{\sigma_{ij}} - 1)e^{\mu_i + \mu_j + \frac{1}{2}(\sigma_{ii} + \sigma_{jj})}, \tag{A.2}$$

$$\tilde{\rho}_{ij} = \frac{e^{\rho\sqrt{\sigma_{ii}\sigma_{jj}}} - 1}{\sqrt{(e^{\sigma_{ii}} - 1)e^{\sigma_{jj}} - 1)}}. \tag{A.3}$$

How this translate into confidence regions is shown in Figure A.1. The figure hopefully give some intuition of how the confidence regions for multidimensional $\log$-normal random variables behave in comparison with the corresponding Gaussian random variable.

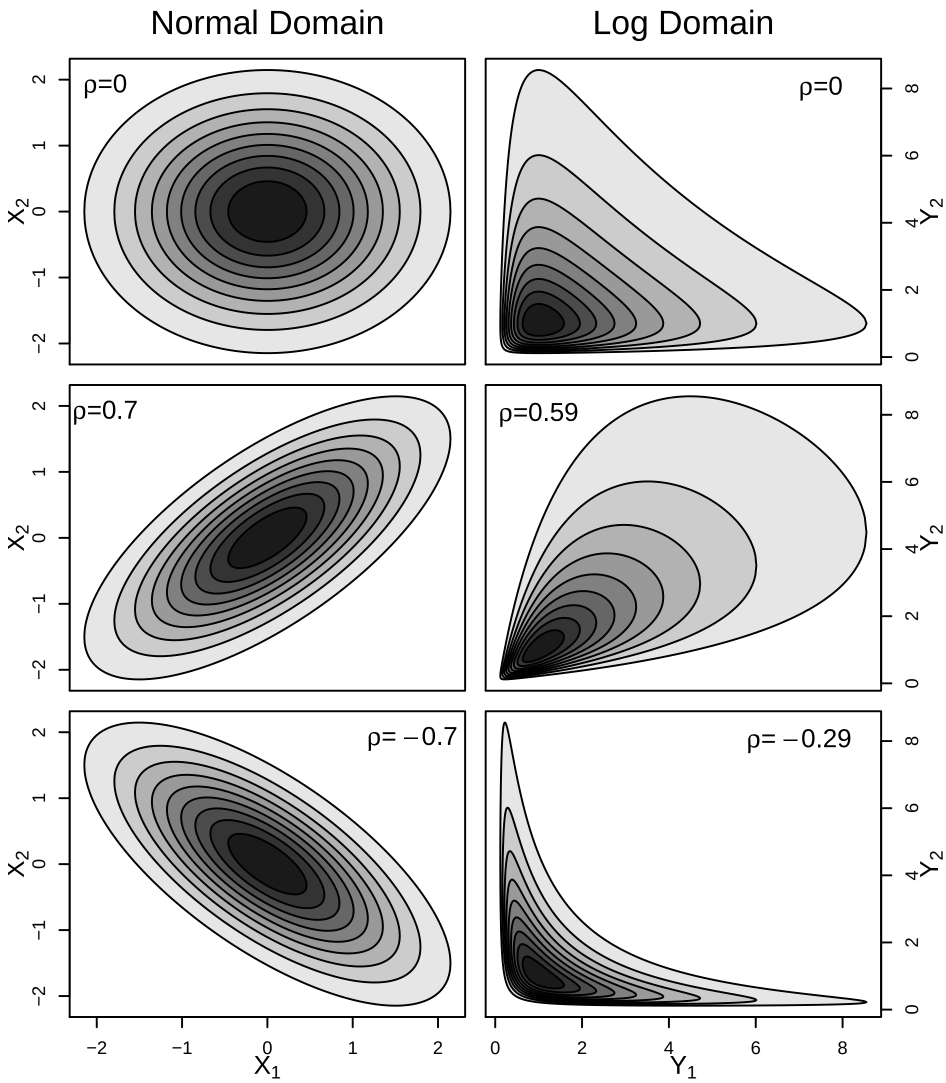


Figure A.1: Gaussian and corresponding $\log$-normal confidence regions, the shown confidence regions are 10% to 90% in steps of 10%. Left column is for Gaussian random variables with $E[X_1] = E[X_2] = 0$ and $V[X_1] = V[X_2] = 1$ and different correlations. Right column is for $Y_i = e^{X_i}$, i.e. the $\log$-normal variable corresponding to the normal variable in the left column, the expectation is $E[Y_1] = E[Y_2] = e^{1/2}$ and variance $V[Y_1] = V[Y_2] = (e-1)e$, the correlations are given in the plots.

# Bibliography


Kotz S., Balakrishnan N. and Johnson L. N. (2000) Continuous Multivariate Distributions, Volume 1: Models and Applications, *Wiley Series in Probability and Statistics*

Limpert E., Stahel W. A. and Abbt Markus (2001) Log-normal Distributions across the Sciences: Keys and Clues, *BioScience*, **51**(5),341-352

Møller J. K., Madsen H. and Carstensen J. (2008) Kalman Filtering and Parameter Estimation in State Space Models with Multiplicative Noise, *Submitted*